

Asymptotics of Constant Step Stochastic Approximations Involving Differential Inclusions

Pascal Bianchi, Walid Hachem, and Adil Salim

LTCI, CNRS, Télécom ParisTech, Université Paris-Saclay, 75013, Paris, France.
 pascal.bianchi,walid.hachem,adil.salim@telecom-paristech.fr

February 15, 2017

Abstract

A constant step size stochastic approximation algorithm involving a Differential Inclusion (DI) is studied. In the considered model, the set valued function defining the DI is the selection integral of an upper semi continuous, convex and closed set valued function. First, the narrow convergence of the interpolated process towards the set of solutions of the DI is established as the step size converges to zero. Second, since the iterates of the algorithm form an homogeneous Markov chain whose transition kernel is indexed by the step size, it is shown that the set of invariant measures of these Markov chains is tight. Using the notion of invariant measures of a set-valued flow recently introduced by Faure and Roth, we prove that the cluster points of these invariant measures are invariant for the set valued flow induced by the DI. By putting the above results together, conclusions in terms of the long run behavior of the iterates in the small step size regime are drawn. The ergodic behavior of the iterates is also addressed.

Keywords: Differential inclusions; Dynamical systems; Narrow convergence of stochastic processes; Stochastic approximation with constant step.
 34A60, 47N10, 54C60, 62L20.

1 Introduction

Consider the iterative algorithm

$$x_{n+1} = x_n + \gamma h_\gamma(\xi_{n+1}, x_n), \quad (1)$$

where (x_n) is a sequence of \mathbb{R}^N -valued random variables, (ξ_n) is a sequence of independent and identically distributed (iid) random variables with probability law μ , γ is a positive step size, and $\{h_\gamma(\xi, x)\}_{\gamma \in (0, \gamma_0)}$ is a family of functions defined for small values of γ . The long term behavior of the iterates (1) in the small step size regime has been studied in the treatises [7, 3, 17, 11] among others. A popular approach is the so-called Ordinary Differential Equation (ODE) method [14, 4], whose principle is as follows. Assume that $h_\gamma(\xi, x)$ converges to a function $H(\xi, x)$ as $\gamma \rightarrow 0$, define the function $H(x) := \mathbb{E}_\xi H(\xi, x)$, and assume that the ODE $\dot{x}(t) = H(x(t))$ has a unique solution on $[0, \infty)$ for any initial condition. Denote by $x_\gamma(t)$ the continuous-time stochastic process obtained by a piecewise linear interpolation of the sequence x_n , where the points x_n are spaced by a fixed time step γ on the positive real axis. The first step of the approach consists in showing that x_γ converges narrowly to the solution of the ODE as $\gamma \rightarrow 0$ (in the topology of uniform convergence on compact sets). The second step consists in establishing a stability result. To that end, the family of processes $\{(x_n)\}_{\gamma \in (0, \gamma_0)}$ is seen as a family of homogeneous Feller Markov chains whose transition kernels are indexed by γ . The idea here is to show that the family of invariant measures of these kernels is tight. Having the ODE convergence and the tightness of the invariant measures,

it is known that the cluster points of these invariant measures as $\gamma \rightarrow 0$ are invariant for the flow generated by the ODE $\dot{x}(t) = H(x(t))$ [16, 14, 4]. In particular, this in turn implies that when γ is small, the process (x_n) spends most of its time in a neighborhood of the Birkhoff center of the flow generated by the ODE.

In this paper, we are interested in the more general case where for μ -almost each ξ and for each sequence $((a_k, \gamma_k))$ converging to some couple $(a, 0)$ as $k \rightarrow \infty$, the sequence $(h_{\gamma_k}(\xi, x_k))$ converges to some *subset* $H(\xi, a)$ of \mathbb{R}^N . In this context, the expectation $\mathbb{E}_\xi H(\xi, x)$ becomes a *selection integral*, which is defined as the closure of the set of integrals of the form $\int \varphi d\mu$ where φ is any integrable function such that $\varphi(\xi) \in H(\xi, x)$ for μ -almost all ξ . Then, the ODE introduced previously must be replaced by a *Differential Inclusion* (DI) $\dot{x}(t) \in H(x(t))$, where $H(x) := \mathbb{E}_\xi H(\xi, x)$.

Our first result consists in establishing the narrow convergence of x_γ towards the set of solutions of the DI in the compact convergence topology, as $\gamma \rightarrow 0$. Here, $H(\xi, \cdot)$ is assumed to be a proper and upper semi continuous (usc) maps with closed convex values. To that end, we show that the family $\{x_\gamma\}$ is tight. Then we characterize the cluster points using weak convergence tools (the Banach-Alaoglu theorem) coupled with the convergence theorem towards usc maps ([1, Chap. 1.4, Th. 1]).

Next, we consider the Markov chains that describe the processes $\{(x_n)\}_{\gamma \in (0, \gamma_0)}$. Assuming that the so-called Pakes-Has'minskii criterion is satisfied, we establish the existence of invariant measures, and the tightness of the family of all invariant measures. The next step is to characterize the cluster points of these invariant measures as $\gamma \rightarrow 0$. Since the flow generated by the DI $\dot{x}(t) \in H(x(t))$ is in general set-valued, the notion of invariant measure is more demanding than in the ODE case. Borrowing from [13] the concept of invariant measure for such flows, we show that the cluster points of the invariant measures are also flow invariant.

Using these results, we obtain that for small γ , most of the iterates stay close to the Birkhoff center of the flow induced by the DI, in a sense made clear in the paper. We also characterize the ergodic behavior of these iterates, showing that when γ is small, the sums $n^{-1} \sum_1^n x_k$ stay close for large n to the convex closure of the limit set of the averaged DI solutions.

The recent interest in the stochastic approximation when the ODE is replaced with a differential inclusion dates back to [5], where decreasing steps were considered. A similar setting is considered in [12]. A Markov noise was considered in the recent manuscript [22]. We also mention [13], where the ergodic convergence is studied when the so called weak asymptotic pseudo trajectory property is satisfied. The case where the DI is built from maximal monotone operators is studied in [8] and [9].

Differential inclusions arise in many applications, which include game theory (see [5, 6], [20] and the references therein), convex optimization [9], queuing theory or wireless communications, where stochastic approximation algorithms with non continuous drifts are frequently used, and can be modelled by differential inclusions [15].

Differential inclusions with a constant step were studied in [20]. We follow here the same general approach as in [20]. However, the stochastic approximation model studied in [20] is different from (1), and reads $x_{k+1} - x_k - \gamma U_{k+1} \in \gamma H_\gamma(x_k)$, where $H_\gamma(x)$ is an approximation of $H(x)$, and where γU_{k+1} is a additive noise whose associated interpolated process converges to zero in probability. Moreover, the ergodic convergence is not considered in [20]. Finally, the proof techniques used in this reference are quite different from ours, and rely heavily on the compactness of the solution set of the DI.

Paper organization. Section 2 is devoted to the mathematical background and to the notations. The main results are given in Section 3. The tightness of the interpolated process as well as its narrow convergence towards the solution set of the DI (Th. 3.1) are proven in Section 4. Turning to the Markov chain characterization, Prop. 3.2, who explores the relations between the cluster points of the Markov chains invariant measures and the invariant measures of the flow induced by

the DI, is proven in Section 5. A general result describing the asymptotic behavior of a functional of the iterates with a prescribed growth is provided by Th. 3.3, and proven in Section 6. Finally, in Section 7, we show how the results pertaining to the ergodic convergence and to the convergence of the iterates (Th. 3.4 and 3.5 respectively) can be deduced from Th. 3.3.

2 Background

2.1 General Notations

The notation $C(E, F)$ is used to denote the set of continuous functions on $E \rightarrow F$. The notation $C_b(E)$ stands for the set of bounded functions in $C(E, \mathbb{R})$. We use the conventions $\sup \emptyset = -\infty$ and $\inf \emptyset = +\infty$. Notation $\lfloor x \rfloor$ stands for the integer part of x .

Let (E, d) be a metric space. For every $x \in E$ and $S \subset E$, we define $d(x, S) = \inf\{d(x, y) : y \in S\}$. We say that a sequence (x_n) on E converges to S , noted $x_n \rightarrow_n S$ or simply $x_n \rightarrow S$, if $d(x_n, S)$ tends to zero as n tends to infinity. The ε -neighborhood of the set S is denoted by $S_\varepsilon := \{x \in E : d(x, S) < \varepsilon\}$. The closure of S is denoted by \overline{S} , its complementary set by S^c and its convex hull by $\text{co}(S)$. The characteristic function of S is the function $1_S : E \rightarrow \{0, 1\}$ equal to one on S and to zero elsewhere.

Let $E = \mathbb{R}^N$ for some integer $N \geq 1$. We endow the space $C(\mathbb{R}_+, E)$ with the topology of uniform convergence on compact sets, where $\mathbb{R}_+ := [0, +\infty)$. The space $C(\mathbb{R}_+, E)$ is metrizable by the distance d defined for every $x, y \in C(\mathbb{R}_+, E)$ by

$$d(x, y) := \sum_{n \in \mathbb{N}^*} 2^{-n} \left(1 \wedge \sup_{t \in [0, n]} \|x(t) - y(t)\| \right), \quad (2)$$

where $\|\cdot\|$ denotes the Euclidean norm in E .

2.2 Random Probability Measures

Let E denote a metric space and let $\mathcal{B}(E)$ be its Borel σ -field. We denote by $\mathcal{M}(E)$ the set of probability measures on $(E, \mathcal{B}(E))$. The support $\text{supp}(\nu)$ of a measure $\nu \in \mathcal{M}(E)$ is the smallest closed set G such that $\nu(G) = 1$. We endow $\mathcal{M}(E)$ with the topology of narrow convergence: a sequence (ν_n) on $\mathcal{M}(E)$ converges to a measure $\nu \in \mathcal{M}(E)$ (noted $\nu_n \Rightarrow \nu$) if for every $f \in C_b(E)$, $\nu_n(f) \rightarrow \nu(f)$. If E is a Polish space, $\mathcal{M}(E)$ is metrizable by the Lévy-Prokhorov distance, and is a Polish space as well. A subset \mathcal{G} of $\mathcal{M}(E)$ is said tight if for every $\varepsilon > 0$, there exists a compact subset K of E such that for all $\nu \in \mathcal{G}$, $\nu(K) > 1 - \varepsilon$. By Prokhorov's theorem, \mathcal{G} is tight if and only if it is relatively compact in $\mathcal{M}(E)$.

We denote by δ_a the Dirac measure at the point $a \in E$. If X is a random variable on some measurable space (Ω, \mathcal{F}) into $(E, \mathcal{B}(E))$, we denote by $\delta_X : \Omega \rightarrow \mathcal{M}(E)$ the measurable mapping defined by $\delta_X(\omega) = \delta_{X(\omega)}$. If $\Lambda : (\Omega, \mathcal{F}) \rightarrow (\mathcal{M}(E), \mathcal{B}(\mathcal{M}(E)))$ is a random variable on the set of probability measures, we denote by $\mathbb{E}\Lambda$ the probability measure defined by $(\mathbb{E}\Lambda)(f) := \mathbb{E}(\Lambda(f))$, for every $f \in C_b(E)$.

2.3 Set-Valued Mappings and Differential Inclusions

A set-valued mapping $H : E \rightrightarrows F$ is a function on E into the set 2^F of subsets of F . The graph of H is $\text{gr}(H) := \{(a, b) \in E \times F : b \in H(a)\}$. The domain of H is $\text{dom}(H) := \{a \in E : H(a) \neq \emptyset\}$. The mapping H is said proper if $\text{dom}(H)$ is non-empty. We say that H is single-valued if $H(a)$ is a singleton for every $a \in E$ (in which case we handle H simply as a function $H : E \rightarrow F$).

Let $H : E \rightrightarrows E$ be a set-valued map on $E = \mathbb{R}^N$, where N is a positive integer. Consider the differential inclusion:

$$\dot{x}(t) \in H(x(t)). \quad (3)$$

We say that an absolutely continuous mapping $x : \mathbb{R}_+ \rightarrow E$ is a solution to the differential inclusion with initial condition $a \in E$ if $x(0) = a$ and if (3) holds for almost every $t \in \mathbb{R}_+$. We denote by

$$\Phi_H : E \rightrightarrows C(\mathbb{R}_+, E)$$

the set-valued mapping such that for every $a \in E$, $\Phi_H(a)$ is set of solutions to (3) with initial condition a . We refer to Φ_H as the evolution system induced by H . For every subset $A \subset E$, we define $\Phi_H(A) = \bigcup_{a \in A} \Phi_H(a)$.

A mapping $H : E \rightrightarrows E$ is said *upper semi continuous* (usc) at a point $a_0 \in E$ if for every open set U containing $H(a_0)$, there exists $\eta > 0$, such that for every $a \in E$, $\|a - a_0\| < \eta$ implies $H(a) \subset U$. It is said usc if it is usc at every point [1, Chap. 1.4]. In the particular case where H is usc with nonempty compact convex values and satisfies the condition

$$\exists c > 0, \forall a \in E, \sup\{\|b\| : b \in H(a)\} \leq c(1 + \|a\|), \quad (4)$$

then, $\text{dom}(\Phi_H) = E$, see e.g. [1], and moreover, $\Phi_H(E)$ is closed in the metric space $(C(\mathbb{R}_+, E), d)$.

2.4 Invariant Measures of Set-Valued Evolution Systems

Let (E, d) be a metric space. We define the shift operator $\Theta : C(\mathbb{R}_+, E) \rightarrow C(\mathbb{R}_+, C(\mathbb{R}_+, E))$ s.t. for every $x \in C(\mathbb{R}_+, E)$, $\Theta(x) : t \mapsto x(t + \cdot)$.

Consider a set-valued mapping $\Phi : E \rightrightarrows C(\mathbb{R}_+, E)$. When Φ is single-valued (i.e., for all $a \in E$, $\Phi(a)$ is a continuous function), a measure $\pi \in \mathcal{M}(E)$ is called an *invariant measure* for Φ , or Φ -invariant, if for all $t > 0$, $\pi = \pi \Phi_t^{-1}$, where $\Phi_t : E \rightarrow E$ is the map defined by $\Phi_t(a) = \Phi(a)(t)$. For all $t \geq 0$, we define the projection $p_t : C(\mathbb{R}_+, E) \rightarrow E$ by $p_t(x) = x(t)$.

The definition can be extended as follows to the case where Φ is set-valued.

Definition 2.1. A probability measure $\pi \in \mathcal{M}(E)$ is said invariant for Φ if there exists $v \in \mathcal{M}(C(\mathbb{R}_+, E))$ s.t.

- (i) $\text{supp}(v) \subset \overline{\Phi(E)}$;
- (ii) v is Θ -invariant;
- (iii) $vp_0^{-1} = \pi$.

When Φ is single valued, both definitions coincide. The above definition is borrowed from [13] (see also [18]). Note that $\overline{\Phi(E)}$ can be replaced by $\Phi(E)$ whenever the latter set is closed (sufficient conditions for this have been provided above).

The limit set of a function $x \in C(\mathbb{R}_+, E)$ is defined as

$$L_x := \bigcap_{t \geq 0} \overline{x([t, +\infty))}.$$

It coincides with the set of points of the form $\lim_n x(t_n)$ for some sequence $t_n \rightarrow \infty$. Consider now a set valued mapping $\Phi : E \rightrightarrows C(\mathbb{R}_+, E)$. The limit set $L_{\Phi(a)}$ of a point $a \in E$ for Φ is

$$L_{\Phi(a)} := \bigcup_{x \in \Phi(a)} L_x,$$

and $L_\Phi := \bigcup_{a \in E} L_{\Phi(a)}$. A point a is said recurrent for Φ if $a \in L_{\Phi(a)}$. The Birkhoff center of Φ is the closure of the set of recurrent points

$$\text{BC}_\Phi := \overline{\{a \in E : a \in L_{\Phi(a)}\}}.$$

The following result, established in [13] (see also [2]), is a consequence of the celebrated recurrence theorem of Poincaré.

Proposition 2.1. Let $\Phi : E \rightrightarrows C(\mathbb{R}_+, E)$. Assume that $\Phi(E)$ is closed. Let $\pi \in \mathcal{M}(E)$ be an invariant measure for Φ . Then, $\pi(\text{BC}_\Phi) = 1$.

We denote by $\mathcal{I}(\Phi)$ the subset of $\mathcal{M}(E)$ formed by all invariant measures for Φ . Finally, we define

$$\mathcal{J}(\Phi) := \{\mathbf{m} \in \mathcal{M}(\mathcal{M}(E)) : \forall A \in \mathcal{B}(\mathcal{M}(E)), \mathcal{I}(\Phi) \subset A \Rightarrow \mathbf{m}(A) = 1\}.$$

Finally, we define the mapping $\text{av} : C(\mathbb{R}_+, E) \rightarrow C(\mathbb{R}_+, E)$ by

$$\text{av}(\mathbf{x}) : t \mapsto \frac{1}{t} \int_0^t \mathbf{x}(s) ds,$$

and $\text{av}(\mathbf{x})(0) = \mathbf{x}(0)$. We define $\text{av}(\Phi) : E \rightrightarrows C(\mathbb{R}_+, E)$ by $\text{av}(\Phi)(a) = \{\text{av}(\mathbf{x}) : \mathbf{x} \in \Phi(a)\}$ for each $a \in E$.

2.5 The Selection Integral

Let (Ξ, \mathcal{G}, μ) denote an arbitrary probability space. For $1 \leq p < \infty$, we denote by $\mathcal{L}^p(\Xi, \mathcal{G}, \mu; E)$ the Banach space of the measurable functions $\varphi : \Xi \rightarrow E$ such that $\int \|\varphi\|^p d\mu < \infty$. For any set-valued mapping $G : \Xi \rightrightarrows E$, we define the set

$$\mathfrak{S}_G^p := \{\varphi \in \mathcal{L}^p(\Xi, \mathcal{G}, \mu; E) : \varphi(\xi) \in G(\xi) \text{ } \mu - \text{a.e.}\}.$$

Any element of \mathfrak{S}_G^1 is referred to as an *integrable selection*. If $\mathfrak{S}_G^1 \neq \emptyset$, the mapping G is said to be integrable. The *selection integral* [19] of G is the set

$$\int G d\mu := \overline{\left\{ \int_{\Xi} \varphi d\mu : \varphi \in \mathfrak{S}_G^1 \right\}}.$$

3 Main Results

3.1 Dynamical Behavior of Stochastic Approximation

From now on to the end of this paper, we set $E := \mathbb{R}^N$ where N is a positive integer. Let (Ξ, \mathcal{G}, μ) be an arbitrary probability space. Choose $\gamma_0 > 0$. For every $\gamma \in (0, \gamma_0)$, let $h_\gamma : \Xi \times E \rightarrow E$ be a $\mathcal{G} \otimes \mathcal{B}(E)/\mathcal{B}(E)$ -measurable map. We introduce the probability transition kernel P_γ on $E \times \mathcal{B}(E) \rightarrow [0, 1]$ such that

$$P_\gamma(a, f) := \int f(x + \gamma h_\gamma(s, a)) \mu(ds) \quad (5)$$

for all $f \in C_b(E)$ and all $a \in E$. Note that P_γ corresponds to the transition kernel of the Markov chain defined by Eq. (1).

Assumption (RM). There exists $H : \Xi \times E \rightrightarrows E$ such that:

- i) For every s μ -a.e. and for every converging sequence $(u_n, \gamma_n) \rightarrow (u^*, 0)$ on $E \times (0, \gamma_0)$,

$$h_{\gamma_n}(s, u_n) \rightarrow H(s, u^*).$$

- ii) For all s μ -a.e., $H(s, \cdot)$ is proper, usc, with closed convex values.
- iii) For every $x \in E$, $H(\cdot, x)$ is μ -integrable. We set $\mathbf{H}(x) := \int H(s, x) \mu(ds)$.
- iv) For every $T > 0$ and every compact set $K \subset E$,

$$\sup\{\|\mathbf{x}(t)\| : t \in [0, T], \mathbf{x} \in \Phi_{\mathbf{H}}(a), a \in K\} < \infty.$$

v) For every compact set $K \subset E$, there exists $\epsilon_K > 0$ such that

$$\sup_{x \in K} \sup_{0 < \gamma < \gamma_0} \int \|h_\gamma(s, x)\|^{1+\epsilon_K} \mu(ds) < \infty. \quad (6)$$

Assumption [iv](#)) requires implicitly that the set of solutions $\Phi_H(a)$ is non-empty for any value of a .

On the canonical space $\Omega := E^{\mathbb{N}}$ equipped with the σ -algebra $\mathcal{F} := \mathcal{B}(E)^{\otimes \mathbb{N}}$, we denote by $X : \Omega \rightarrow E^{\mathbb{N}}$ the canonical process defined by $X_n(\omega) = \omega_n$ for every $\omega = (\omega_k : k \in \mathbb{N})$ and every $n \in \mathbb{N}$, where $X_n(\omega)$ is the n -th coordinate of $X(\omega)$. For every $\nu \in \mathcal{M}(E)$ and $\gamma \in (0, \gamma_0)$, we denote by $\mathbb{P}^{\nu, \gamma}$ the unique probability measure on (Ω, \mathcal{F}) such that X is an homogeneous Markov chain with initial distribution ν and transition kernel P_γ . We denote by $\mathbb{E}^{\nu, \gamma}$ the corresponding expectation. When $\nu = \delta_a$ for some $a \in E$, we shall prefer the notation $\mathbb{P}^{a, \gamma}$ to $\mathbb{P}^{\delta_a, \gamma}$.

The set $C(\mathbb{R}_+, E)$ is equipped with the topology of uniform convergence on the compact intervals, who is known to be compatible with the distance d defined by [\(2\)](#). For every $\gamma > 0$, we introduce the measurable map on $(\Omega, \mathcal{F}) \rightarrow (C(\mathbb{R}_+, E), \mathcal{B}(C(\mathbb{R}_+, E)))$, such that for every $x = (x_n : n \in \mathbb{N})$ in Ω ,

$$X_\gamma(x) : t \mapsto x_{\lfloor \frac{t}{\gamma} \rfloor} + (t/\gamma - \lfloor t/\gamma \rfloor)(x_{\lfloor \frac{t}{\gamma} \rfloor + 1} - x_{\lfloor \frac{t}{\gamma} \rfloor}).$$

The random variable X_γ will be referred to as the linearly *interpolated process*. On the space $(C(\mathbb{R}_+, E), \mathcal{B}(C(\mathbb{R}_+, E)))$, the distribution of the r.v. X_γ is $\mathbb{P}^{\nu, \gamma} X_\gamma^{-1}$.

Theorem 3.1. Suppose that Assumption (RM) is satisfied. Then, for every compact set $K \subset E$, the family $\{\mathbb{P}^{a, \gamma} X_\gamma^{-1} : a \in K, 0 < \gamma < \gamma_0\}$ is tight. Moreover, for every $\varepsilon > 0$,

$$\sup_{a \in K} \mathbb{P}^{a, \gamma} (d(X_\gamma, \Phi_H(K)) > \varepsilon) \xrightarrow{\gamma \rightarrow 0} 0.$$

3.2 Convergence Analysis

For each $\gamma \in (0, \gamma_0)$, we denote by

$$\mathcal{I}(P_\gamma) := \{\pi \in \mathcal{M}(E) : \pi = \pi P_\gamma\}$$

the set of invariant probability measures of P_γ . Letting $\mathcal{P} = \{P_\gamma : 0 < \gamma < \gamma_0\}$, we define $\mathcal{I}(\mathcal{P}) = \bigcup_{\gamma \in (0, \gamma_0)} \mathcal{I}(P_\gamma)$. We say that a measure $\nu \in \mathcal{M}(E)$ is a cluster point of $\mathcal{I}(\mathcal{P})$ as $\gamma \rightarrow 0$, if there exists a sequence $\gamma_j \rightarrow 0$ and a sequence of measures $(\pi_j : j \in \mathbb{N})$ s.t. $\pi_j \in \mathcal{I}(P_{\gamma_j})$ for all j , and $\pi_j \Rightarrow \nu$.

We define

$$\mathcal{J}(P_\gamma) := \{\mathbf{m} \in \mathcal{M}(\mathcal{M}(E)) : \text{supp}(\mathbf{m}) \subset \mathcal{I}(P_\gamma)\},$$

and $\mathcal{J}(\mathcal{P}) = \bigcup_{\gamma \in (0, \gamma_0)} \mathcal{J}(P_\gamma)$. We say that a measure $\mathbf{m} \in \mathcal{M}(\mathcal{M}(E))$ is a cluster point of $\mathcal{J}(\mathcal{P})$ as $\gamma \rightarrow 0$, if there exists a sequence $\gamma_j \rightarrow 0$ and a sequence of measures $(\mathbf{m}_j : j \in \mathbb{N})$ s.t. $\mathbf{m}_j \in \mathcal{J}(P_{\gamma_j})$ for all j , and $\mathbf{m}_j \Rightarrow \mathbf{m}$.

Proposition 3.2. Suppose that Assumption (RM) is satisfied. Then,

- i) As $\gamma \rightarrow 0$, any cluster point of $\mathcal{I}(\mathcal{P})$ is an element of $\mathcal{I}(\Phi_H)$;
- ii) As $\gamma \rightarrow 0$, any cluster point of $\mathcal{J}(\mathcal{P})$ is an element of $\mathcal{J}(\Phi_H)$.

In order to explore the consequences of this proposition, we introduce two supplementary assumptions. The first is the so-called Pakes-Has'minskii tightness criterion, who reads as follows [\[14\]](#):

Assumption (PH). There exists measurable mappings $V : E \rightarrow [0, +\infty)$, $\psi : E \rightarrow [0, +\infty)$ s.t. $\psi(x) \rightarrow +\infty$ as $\|x\| \rightarrow +\infty$, and two functions $\alpha : (0, \gamma_0) \rightarrow (0, +\infty)$, $\beta : (0, \gamma_0) \rightarrow \mathbb{R}$, such that

$$\sup_{\gamma \in (0, \gamma_0)} \frac{\beta(\gamma)}{\alpha(\gamma)} < \infty \quad \text{and} \quad \lim_{\|x\| \rightarrow +\infty} \psi(x) = +\infty,$$

and for every $\gamma \in (0, \gamma_0)$,

$$P_\gamma V \leq V - \alpha(\gamma)\psi + \beta(\gamma).$$

We recall that a transition kernel P on $E \times \mathcal{B}(E) \rightarrow [0, 1]$ is said *Feller* if the mapping $Pf : x \mapsto \int f(y)P(x, dy)$ is continuous for any $f \in C_b(E)$. If P is Feller, then the set of invariant measures of P is a closed subset of $\mathcal{M}(E)$. The following assumption ensures that for all $\gamma \in (0, \gamma_0)$, P_γ is Feller.

Assumption (FL). For every $s \in \Xi$, $\gamma \in (0, \gamma_0)$, the function $h_\gamma(s, \cdot)$ is continuous.

Theorem 3.3. Let Assumptions (RM), (PH) and (FL) be satisfied. Let ψ and V be the functions specified in (PH). Let $\nu \in \mathcal{M}(E)$ s.t. $\nu(V) < \infty$. Let $\mathcal{U} := \bigcup_{\pi \in \mathcal{I}(\Phi)} \text{supp}(\pi)$. Then, for all $\varepsilon > 0$,

$$\limsup_{n \rightarrow \infty} \frac{1}{n+1} \sum_{k=0}^n \mathbb{P}^{\nu, \gamma}(d(X_k, \mathcal{U}) > \varepsilon) \xrightarrow{\gamma \rightarrow 0} 0. \quad (7)$$

Let $N' \in \mathbb{N}^*$ and $f \in C(E, \mathbb{R}^{N'})$. Assume that there exists $M \geq 0$ and $\varphi : \mathbb{R}^{N'} \rightarrow \mathbb{R}_+$ such that $\lim_{\|a\| \rightarrow \infty} \varphi(a)/\|a\| = \infty$ and

$$\forall a \in E, \quad \varphi(f(a)) \leq M(1 + \psi(a)). \quad (8)$$

Then, the set $\mathcal{S}_f := \{\pi(f) : \pi \in \mathcal{I}(\Phi) \text{ and } \pi(\|f(\cdot)\|) < \infty\}$ is nonempty. For all $n \in \mathbb{N}$, $\gamma \in (0, \gamma_0)$, the r.v.

$$F_n := \frac{1}{n+1} \sum_{k=0}^n f(X_k)$$

is $\mathbb{P}^{\nu, \gamma}$ -integrable, and satisfies for all $\varepsilon > 0$,

$$\limsup_{n \rightarrow \infty} d(\mathbb{E}^{\nu, \gamma}(F_n), \mathcal{S}_f) \xrightarrow{\gamma \rightarrow 0} 0, \quad (9)$$

$$\limsup_{n \rightarrow \infty} \mathbb{P}^{\nu, \gamma}(d(F_n, \mathcal{S}_f) \geq \varepsilon) \xrightarrow{\gamma \rightarrow 0} 0. \quad (10)$$

Theorem 3.4. Let Assumptions (RM), (PH) and (FL) be satisfied. Assume that $\Phi_H(E)$ is closed. Let ψ and V be the functions specified in (PH). Let $\nu \in \mathcal{M}(E)$ s.t. $\nu(V) < \infty$. Assume that

$$\lim_{\|a\| \rightarrow \infty} \frac{\psi(a)}{\|a\|} = +\infty.$$

For all $n \in \mathbb{N}$, define $\overline{X}_n := \frac{1}{n+1} \sum_{k=0}^n X_k$. Then, for all $\varepsilon > 0$,

$$\limsup_{n \rightarrow \infty} d(\mathbb{E}^{\nu, \gamma}(\overline{X}_n), \text{co}(L_{\text{av}}(\Phi))) \xrightarrow{\gamma \rightarrow 0} 0,$$

$$\limsup_{n \rightarrow \infty} \mathbb{P}^{\nu, \gamma}(d(\overline{X}_n, \text{co}(L_{\text{av}}(\Phi))) \geq \varepsilon) \xrightarrow{\gamma \rightarrow 0} 0.$$

Theorem 3.5. Let Assumptions (RM), (PH) and (FL) be satisfied. Assume that $\Phi_H(E)$ is closed. Let ψ and V be the functions specified in (PH). Let $\nu \in \mathcal{M}(E)$ s.t. $\nu(V) < \infty$. Then, for all $\varepsilon > 0$,

$$\limsup_{n \rightarrow \infty} \frac{1}{n+1} \sum_{k=0}^n \mathbb{P}^{\nu, \gamma}(d(X_k, \text{BC}_\Phi) \geq \varepsilon) \xrightarrow{\gamma \rightarrow 0} 0.$$

4 Proof of Theorem 3.1

The first lemma is a straightforward adaptation of the *convergence theorem* [1, Chap. 1.4, Th. 1, pp. 60]. Hence, the proof is omitted. We denote by λ_T the Lebesgue measure on $[0, T]$.

Lemma 4.1. Let $\{F_\xi : \xi \in \Xi\}$ be a family of mappings on $E \rightrightarrows E$. Let $T > 0$ and for all $n \in \mathbb{N}$, let $u_n : [0, T] \rightarrow E$, $v_n : \Xi \times [0, T] \rightarrow E$ be measurable maps w.r.t $\mathcal{B}([0, T])$ and $\mathcal{G} \otimes \mathcal{B}([0, T])$ respectively. Note for simplicity $\mathcal{L}^1 := \mathcal{L}^1(\Xi \times [0, T], \mathcal{G} \otimes \mathcal{B}([0, T]), \mu \otimes \lambda_T; \mathbb{R})$. Assume the following.

- i) For all (ξ, t) $\mu \otimes \lambda_T$ -a.e., $(u_n(t), v_n(\xi, t)) \rightarrow_n \text{gr}(F_\xi)$.
- ii) (u_n) converges λ_T -a.e. to a function $u : [0, T] \rightarrow E$.
- iii) For all n , $v_n \in \mathcal{L}^1$ and converges weakly in \mathcal{L}^1 to a function $v : \Xi \times [0, T] \rightarrow E$.
- iv) For all ξ μ -a.e., F_ξ is proper upper semi continuous with closed convex values.

Then, for all (ξ, t) $\mu \otimes \lambda_T$ -a.e., $v(\xi, t) \in F_\xi(u(t))$.

Given $T > 0$ and $0 < \delta \leq T$, we denote by

$$w_x^T(\delta) := \sup\{\|\mathbf{x}(t) - \mathbf{x}(s)\| : |t - s| \leq \delta, (t, s) \in [0, T]^2\}$$

the modulus of continuity on $[0, T]$ of any $\mathbf{x} \in C(\mathbb{R}_+, E)$.

Lemma 4.2. For all $n \in \mathbb{N}$, denote by $\mathcal{F}_n \subset \mathcal{F}$ the σ -field generated by the r.v. $\{X_k : 0 \leq k \leq n\}$. For all $\gamma \in (0, \gamma_0)$, define $Z_{n+1}^\gamma := \gamma^{-1}(X_{n+1} - X_n)$. Let $K \subset E$ be compact. Let $\{\bar{\mathbb{P}}^{a, \gamma} : a \in K, 0 < \gamma < \gamma_0\}$ be a family of probability measures on (Ω, \mathcal{F}) satisfying the following uniform integrability condition:

$$\sup_{n \in \mathbb{N}^*, a \in K, \gamma \in (0, \gamma_0)} \bar{\mathbb{E}}^{a, \gamma}(\|Z_n^\gamma\| \mathbb{1}_{\|Z_n^\gamma\| > A}) \xrightarrow{A \rightarrow +\infty} 0. \quad (11)$$

Then, $\{\bar{\mathbb{P}}^{a, \gamma} X_\gamma^{-1} : a \in K, 0 < \gamma < \gamma_0\}$ is tight. Moreover, for any $T > 0$,

$$\sup_{a \in K} \bar{\mathbb{P}}^{a, \gamma} \left(\max_{0 \leq n \leq \lfloor \frac{T}{\gamma} \rfloor} \gamma \left\| \sum_{k=0}^n (Z_{k+1}^\gamma - \bar{\mathbb{E}}^{a, \gamma}(Z_{k+1}^\gamma | \mathcal{F}_k)) \right\| > \varepsilon \right) \xrightarrow{\gamma \rightarrow 0} 0. \quad (12)$$

Proof. We prove the first point. Set $T > 0$, let $0 < \delta \leq T$, and choose $0 \leq s \leq t \leq T$ s.t. $t - s \leq \delta$. Let $\gamma \in (0, \gamma_0)$ and set $n := \lfloor \frac{t}{\gamma} \rfloor$, $m := \lfloor \frac{s}{\gamma} \rfloor$. For any $R > 0$,

$$\|X_\gamma(t) - X_\gamma(s)\| \leq \gamma \sum_{k=m+1}^{n+1} \|Z_k^\gamma\| \leq \gamma(n - m + 1)R + \gamma \sum_{k=m+1}^{n+1} \|Z_k^\gamma\| \mathbb{1}_{\|Z_k^\gamma\| > R}.$$

Noting that $n - m + 1 \leq \frac{\delta}{\gamma}$ and using Markov inequality, we obtain

$$\begin{aligned} \bar{\mathbb{P}}^{a, \gamma} X_\gamma^{-1}(\{x : w_x^T(\delta) > \varepsilon\}) &\leq \bar{\mathbb{P}}^{a, \gamma} \left(\gamma \sum_{k=1}^{\lfloor \frac{T}{\gamma} \rfloor + 1} \|Z_k^\gamma\| \mathbb{1}_{\|Z_k^\gamma\| > R} > \varepsilon - \delta R \right) \\ &\leq T \frac{\sup_{k \in \mathbb{N}^*} \bar{\mathbb{E}}^{a, \gamma}(\|Z_k^\gamma\| \mathbb{1}_{\|Z_k^\gamma\| > R})}{\varepsilon - \delta R}, \end{aligned}$$

provided that $R\delta < \varepsilon$. Choosing $R = \varepsilon/(2\delta)$ and using the uniform integrability,

$$\sup_{a \in K, 0 < \gamma < \gamma_0} \bar{\mathbb{P}}^{a, \gamma} X_\gamma^{-1}(\{x : w_x^T(\delta) > \varepsilon\}) \xrightarrow{\delta \rightarrow 0} 0.$$

As $\{\bar{\mathbb{P}}^{a,\gamma} \mathbf{X}_\gamma^{-1} p_0^{-1} : 0 < \gamma < \gamma_0, a \in K\}$ is obviously tight, the tightness of $\{\bar{P}^{a,\gamma} \mathbf{X}_\gamma^{-1} : a \in K, 0 < \gamma < \gamma_0\}$ follows from [10, Theorem 7.3].

We prove the second point. We define $M_{n+1}^{a,\gamma} := \sum_{k=0}^n (Z_{k+1}^\gamma - \bar{\mathbb{E}}^{a,\gamma}(Z_{k+1}^\gamma | \mathcal{F}_k))$. We introduce

$$\eta_{n+1}^{a,\gamma,\leq} := Z_{n+1}^\gamma \mathbb{1}_{\|Z_{n+1}^\gamma\| \leq R} - \bar{\mathbb{E}}^{a,\gamma}(Z_{n+1}^\gamma \mathbb{1}_{\|Z_{n+1}^\gamma\| \leq R} | \mathcal{F}_n)$$

and we define $\eta_{n+1}^{a,\gamma,>}$ in a similar way, by replacing \leq with $>$ in the right hand side of the above equation. Clearly, for all $a \in K$, $M_{n+1}^{a,\gamma} = \eta_{n+1}^{a,\gamma,\leq} + \eta_{n+1}^{a,\gamma,>}$. Thus,

$$\gamma \|M_{n+1}^{a,\gamma}\| \leq \|S_{n+1}^{a,\gamma,\leq}\| + \|S_{n+1}^{a,\gamma,>}\|$$

where $S_{n+1}^{a,\gamma,\leq} := \gamma \sum_{k=0}^n \eta_{k+1}^{a,\gamma,\leq}$ and $S_{n+1}^{a,\gamma,>}$ is defined similarly. Under $\bar{\mathbb{P}}^{a,\gamma}$, the random processes $S^{a,\gamma,\leq}$ and $S^{a,\gamma,>}$ are \mathcal{F}_n -adapted martingales. Defining $q_\gamma := \lfloor \frac{T}{\gamma} \rfloor + 1$, we obtain by Doob's martingale inequality and by the boundedness of the increments of $S_n^{a,\gamma,\leq}$ that

$$\bar{\mathbb{P}}^{a,\gamma} \left(\max_{1 \leq n \leq q_\gamma} \|S_n^{a,\gamma,\leq}\| > \varepsilon \right) \leq \frac{\bar{\mathbb{E}}^{a,\gamma}(\|S_{q_\gamma}^{a,\gamma,\leq}\|)}{\varepsilon} \leq \frac{\bar{\mathbb{E}}^{a,\gamma}(\|S_{q_\gamma}^{a,\gamma,\leq}\|^2)^{1/2}}{\varepsilon} \leq \frac{2}{\varepsilon} \gamma R \sqrt{q_\gamma},$$

and the right hand side tends to zero uniformly in $a \in K$ as $\gamma \rightarrow 0$. By the same inequality,

$$\bar{\mathbb{P}}^{a,\gamma} \left(\max_{1 \leq n \leq q_\gamma} \|S_n^{a,\gamma,>}\| > \varepsilon \right) \leq \frac{2}{\varepsilon} q_\gamma \gamma \sup_{k \in \mathbb{N}^*} \bar{\mathbb{E}}^{a,\gamma}(\|Z_k^\gamma\| \mathbb{1}_{\|Z_k^\gamma\| > R}).$$

Choose an arbitrarily small $\delta > 0$ and select R as large as need in order that the supremum in the right hand side is no larger than $\varepsilon \delta / (2T + 2\gamma_0)$. Then the left hand side is no larger than δ . Hence, the proof is concluded. \square

For any $R > 0$, define $h_{\gamma,R}(s, a) := h_\gamma(s, a) \mathbb{1}_{\|a\| \leq R}$. Let $H_R(s, x) := H(s, x)$ if $\|x\| < R$, $\{0\}$ if $\|x\| > R$, and E otherwise. Denote the corresponding selection integral as $\mathbf{H}_R(a) = \int H_R(s, a) \mu(ds)$. Define $\tau_R(x) := \inf\{n \in \mathbb{N} : x_n > R\}$ for all $x \in \Omega$. We also introduce the measurable mapping $B_R : \Omega \rightarrow \Omega$, given by

$$B_R(x) : n \mapsto x_n \mathbb{1}_{n < \tau_R(x)} + x_{\tau_R(x)} \mathbb{1}_{n \geq \tau_R(x)}$$

for all $x \in \Omega$ and all $n \in \mathbb{N}$.

Lemma 4.3. Suppose that Assumption (RM) is satisfied. Then, for every compact set $K \subset E$, the family $\{\mathbb{P}^{a,\gamma} B_R^{-1} \mathbf{X}_\gamma^{-1} : \gamma \in (0, \gamma_0), a \in K\}$ is tight. Moreover, for every $\varepsilon > 0$,

$$\sup_{a \in K} \mathbb{P}^{a,\gamma} B_R^{-1} [d(\mathbf{X}_\gamma, \Phi_{\mathbf{H}_R}(K)) > \varepsilon] \xrightarrow{\gamma \rightarrow 0} 0.$$

Proof. We introduce the measurable mapping $\Delta_{\gamma,R} : \Omega \rightarrow E^{\mathbb{N}}$ s.t. for all $x \in \Omega$, $\Delta_{\gamma,R}(x)(0) := 0$ and

$$\Delta_{\gamma,R}(x)(n) := \frac{x_n - x_{n-1}}{\gamma} - \int h_{\gamma,R}(s, x_n) \mu(ds)$$

for all $n \in \mathbb{N}^*$. We also introduce the measurable mapping $\mathbf{G}_{\gamma,R} : C(\mathbb{R}_+, E) \rightarrow C(\mathbb{R}_+, E)$ s.t. for all $x \in C(\mathbb{R}_+, E)$,

$$\mathbf{G}_{\gamma,R}(x)(t) := \int_0^t \int h_{\gamma,R}(s, x(\gamma \lfloor s/\gamma \rfloor)) \mu(ds).$$

We first express the interpolated process in integral form. For every $x \in E^{\mathbb{N}}$ and $t \geq 0$,

$$\mathbf{X}_\gamma(x)(t) = x_0 + \int_0^t \gamma^{-1} (x_{\lfloor \frac{u}{\gamma} \rfloor + 1} - x_{\lfloor \frac{u}{\gamma} \rfloor}) du,$$

from which we obtain the decomposition

$$\mathbf{X}_\gamma(x) = x_0 + \mathbf{G}_{\gamma,R} \circ \mathbf{X}_\gamma(x) + \mathbf{X}_\gamma \circ \Delta_{\gamma,R}(x). \quad (13)$$

The uniform integrability condition (11) is satisfied when letting $\bar{\mathbb{P}}^{a,\gamma} := \mathbb{P}^{a,\gamma} B_R^{-1}$. Indeed,

$$\begin{aligned} \bar{\mathbb{E}}^{a,\gamma}(\|\gamma^{-1}(X_{n+1} - X_n)\|^{1+\epsilon_K}) &= \mathbb{E}^{a,\gamma}(\|\gamma^{-1}(X_{n+1} - X_n)\|^{1+\epsilon_K} \mathbb{1}_{\tau_R(X) > n}) \\ &\leq \sup_{a: \|a\| \leq R} \int \|h_\gamma(s, a)\|^{1+\epsilon_K} \mu(ds), \end{aligned}$$

and the condition (11) follows from hypothesis (6). On the other hand, it is straightforward to show that for all $x \in \Omega$, $\bar{\mathbb{E}}^{a,\gamma}(\gamma^{-1}(X_{n+1} - X_n) | \mathcal{F}_n) = \int h_{\gamma,R}(s, X_n) \mu(ds)$. Thus, Lemma 4.2 implies that for every $\varepsilon > 0$ and all $T > 0$,

$$\sup_{a \in K} \bar{\mathbb{P}}^{a,\gamma} \left(\max_{0 \leq n \leq \lfloor \frac{T}{\gamma} \rfloor} \gamma \left\| \sum_{k=0}^n (\Delta_{\gamma,R})_{k+1} \right\| > \varepsilon \right) \xrightarrow{\gamma \rightarrow 0} 0.$$

It is easy to see that for all $x \in \Omega$, the function $\mathbf{X}_\gamma \circ \Delta_{\gamma,R}(x)$ is bounded on every compact interval $[0, T]$ by $\max_{n \leq \lfloor T/\gamma \rfloor} \|\sum_{k \leq n} \Delta_{k+1}^\gamma\|$. This in turns leads to:

$$\sup_{a \in K} \bar{\mathbb{P}}^{a,\gamma}(\|\mathbf{X}_\gamma \circ \Delta_{\gamma,R}\|_{\infty, T} > \varepsilon) \xrightarrow{\gamma \rightarrow 0} 0, \quad (14)$$

where the notation $\|\mathbf{x}\|_{\infty, T}$ stands for the uniform norm of \mathbf{x} on $[0, T]$.

As a second consequence of Lemma 4.2, the family $\{\bar{\mathbb{P}}^{a,\gamma} \mathbf{X}_\gamma^{-1} : 0 < \gamma < \gamma_0, a \in K\}$ is tight. Choose any subsequence (a_n, γ_n) s.t. $\gamma_n \rightarrow 0$ and $a_n \in K$. Using Prokhorov's theorem and the compactness of K , there exists a subsequence (which we still denote by (a_n, γ_n)) and there exist some $a^* \in K$ and some $v \in \mathcal{M}(C(\mathbb{R}_+, E))$ such that $a_n \rightarrow a^*$ and $\bar{\mathbb{P}}^{a_n, \gamma_n} \mathbf{X}_{\gamma_n}^{-1}$ converges narrowly to v . By Skorokhod's representation theorem, we introduce some r.v. $\mathbf{z}, \{\mathbf{x}_n : n \in \mathbb{N}\}$ on $C(\mathbb{R}_+, E)$ with respective distributions v and $\bar{\mathbb{P}}^{a_n, \gamma_n} \mathbf{X}_{\gamma_n}^{-1}$, defined on some other probability space $(\Omega', \mathcal{F}', \mathbb{P}')$ and such that $d(\mathbf{x}_n(\omega), \mathbf{z}(\omega)) \rightarrow 0$ for all $\omega \in \Omega'$. By (13) and (14), the sequence of r.v.

$$\bar{\mathbf{x}}_n - \bar{\mathbf{x}}_n(0) - \mathbf{G}_{\gamma_n, R}(\mathbf{x}_n)$$

converges in probability to zero in $(\Omega', \mathcal{F}', \mathbb{P}')$, as $n \rightarrow \infty$. One can extract a subsequence under which this convergence holds in the almost sure sense. Therefore, there exists an event of probability one s.t., everywhere on this event,

$$\mathbf{z}(t) = \mathbf{z}(0) + \lim_{n \rightarrow \infty} \int_0^t \int_{\Xi} h_{\gamma_n, R}(s, \mathbf{x}_n(\gamma_n \lfloor u/\gamma_n \rfloor)) \mu(ds) du \quad (\forall t \geq 0),$$

where the limit is of course taken along the former subsequence. We now select an ω s.t. the above convergence holds, and omit the dependence on ω in the sequel (otherwise stated, \mathbf{z} and \mathbf{x}_n are treated as elements of $C(\mathbb{R}_+, E)$ and no longer as random variables). Set $T > 0$. As \mathbf{x}_n converges uniformly on $[0, T]$, there exists a compact set K' (which depends on ω) such that $\mathbf{x}_n(\gamma_n \lfloor t/\gamma_n \rfloor) \in K'$ for all $t \in [0, T]$, $n \in \mathbb{N}$. Define

$$v_n(s, t) := h_{\gamma_n, R}(s, \mathbf{x}_n(\gamma_n \lfloor t/\gamma_n \rfloor)).$$

By Eq. (6), the sequence $(v_n : n \in \mathbb{N})$ forms a bounded subset of $\mathcal{L}^{1+\epsilon_{K'}} := \mathcal{L}^{1+\epsilon_{K'}}(\Xi \times [0, T], \mathcal{G} \otimes \mathcal{B}([0, T]), \mu \otimes \lambda_T; E)$. By the Banach-Alaoglu theorem, the sequence converges weakly to some mapping $v \in \mathcal{L}^{1+\epsilon_{K'}}$ along some subsequence. This has two consequences. First,

$$\mathbf{z}(t) = \mathbf{z}(0) + \int_0^t \int_{\Xi} v(s, u) \mu(ds) du, \quad (\forall t \in [0, T]). \quad (15)$$

Second, for $\mu \otimes \lambda_T$ -almost all (s, t) , $v(s, t) \in H_R(s, z(t))$. In order to prove this point, remark that, by Assumption (RM),

$$v_n(s, t) \rightarrow H_R(s, z(t))$$

for almost all (s, t) . This implies that the couple $(x_n(\gamma_n \lfloor t/\gamma_n \rfloor), v_n(s, t))$ converges to $\text{gr}(H_R(s, \cdot))$ and the second point thus follows from Lemma 4.1. By Fubini's theorem, there exists a negligible set of $[0, T]$ s.t. for all t outside this set, $v(\cdot, t)$ is an integrable selection of $H_R(\cdot, z(t))$. As $H(\cdot, x)$ is integrable for every $x \in E$, the same holds for H_R . Equation (15) implies that $z \in \Phi_{H_R}(K)$. We have shown that for any sequence (a_n, γ_n) on $K \times (0, \gamma_0)$ s.t. $\gamma_n \rightarrow 0$, there exists a subsequence along which, for every $\varepsilon > 0$, $\mathbb{P}^{a_n, \gamma_n} B_R^{-1}(d(X_{\gamma_n}, \Phi_{H_R}(K))) > \varepsilon \rightarrow 0$. This proves the lemma. \square

End of the proof of Theorem 3.1.

We first prove the second statement. Set an arbitrary $T > 0$. Define $d_T(x, y) := \|x - y\|_{\infty, T}$. It is sufficient to prove that for any sequence (a_n, γ_n) s.t. $\gamma_n \rightarrow 0$, there exists a subsequence along which $\mathbb{P}^{a_n, \gamma_n}(d_T(X_{\gamma_n} \circ B_R, \Phi_H(K)) > \varepsilon) \rightarrow 0$. Choose $R > R_0(T)$, where $R_0(T) := \sup\{\|x(t)\| : t \in [0, T], x \in \Phi_H(a), a \in K\}$ is finite by Assumption (RM). It is easy to show that any $x \in \Phi_{H_R}(K)$ must satisfy $\|x\|_{\infty, T} < R$. Thus, when $R > R_0(T)$, any $x \in \Phi_{H_R}(K)$ is such that there exists $y \in \Phi_H(K)$ with $d_T(x, y) = 0$ i.e., the restrictions of x and y to $[0, T]$ coincide. As a consequence of the Lemma 4.3, each sequence (a_n, γ_n) chosen as above admits a subsequence along which, for all $\varepsilon > 0$,

$$\mathbb{P}^{a_n, \gamma_n}(d_T(X_{\gamma_n} \circ B_R, \Phi_H(K)) > \varepsilon) \rightarrow 0. \quad (16)$$

The event $d_T(X_{\gamma_n} \circ B_R, X_{\gamma_n}) > 0$ implies the event $\|X_{\gamma_n} \circ B_R\|_{\infty, T} \geq R$, which in turn implies by the triangular inequality that $R_0(T) + d_T(X_{\gamma_n} \circ B_R, \Phi_H(K)) \geq R$. Therefore,

$$\mathbb{P}^{a_n, \gamma_n}(d_T(X_{\gamma_n} \circ B_R, X_{\gamma_n}) > \varepsilon) \leq \mathbb{P}(d_T(X_{\gamma_n} \circ B_R, \Phi_H(K)) \geq R - R_0(T)). \quad (17)$$

By (16), the right hand side converges to zero. Using (16) again, and the triangular inequality, it follows that $\mathbb{P}^{a_n, \gamma_n}(d_T(X_{\gamma_n}, \Phi_H(K)) > \varepsilon) \rightarrow 0$, which proves the second statement of the theorem.

We prove the first statement (tightness). By [10, Theorem 7.3], this is equivalent to showing that for every $T > 0$, and for every sequence (a_n, γ_n) on $K \times (0, \gamma_0)$, the sequence $(\mathbb{P}^{a_n, \gamma_n} X_{\gamma_n}^{-1} p_0^{-1})$ is tight, and for each positive ε and η , there exists $\delta > 0$ such that $\limsup_n \mathbb{P}^{a_n, \gamma_n} X_{\gamma_n}^{-1}(\{x : w_x^T(\delta) > \varepsilon\}) < \eta$.

First consider the case where $\gamma_n \rightarrow 0$. Fixing $T > 0$, letting $R > R_0(T)$ and using (17), it holds that for all $\varepsilon > 0$, $\mathbb{P}^{a_n, \gamma_n}(d_T(X_{\gamma_n} \circ B_R, X_{\gamma_n}) > \varepsilon) \rightarrow_n 0$. Moreover, we showed that $\mathbb{P}^{a_n, \gamma_n} B_R^{-1} X_{\gamma_n}^{-1}$ is tight. The tightness of $(\mathbb{P}^{a_n, \gamma_n} X_{\gamma_n}^{-1} p_0^{-1})$ follows. In addition, for every $x, y \in C(\mathbb{R}_+, E)$, it holds by the triangle inequality that $w_x^T(\delta) \leq w_y^T(\delta) + 2d_T(x, y)$ for every $\delta > 0$. Thus,

$$\begin{aligned} \mathbb{P}^{a_n, \gamma_n} X_{\gamma_n}^{-1}(\{x : w_x^T(\delta) > \varepsilon\}) &\leq \mathbb{P}^{a_n, \gamma_n} B_R^{-1} X_{\gamma_n}^{-1}(\{x : w_x^T(\delta) > \varepsilon/2\}) \\ &\quad + \mathbb{P}^{a_n, \gamma_n}(d_T(X_{\gamma_n} \circ B_R, X_{\gamma_n}) > \varepsilon/4), \end{aligned}$$

which leads to the tightness of $(\mathbb{P}^{a_n, \gamma_n} X_{\gamma_n}^{-1})$ when $\gamma_n \rightarrow 0$.

It remains to establish the tightness when $\liminf_n \gamma_n > \eta > 0$ for some $\eta > 0$. Note that for all $\gamma > \eta$,

$$w_{X_\gamma^T(x)}(\delta) \leq 2\delta \max_{k=0 \dots \lfloor T/\eta \rfloor + 1} \|x_k\|.$$

There exist n_0 such that for all $n \geq n_0$, $\gamma_n > \eta$ which implies by the union bound:

$$\mathbb{P}^{a_n, \gamma_n} X_{\gamma_n}^{-1}(\{x : w_x^T(\delta) > \varepsilon\}) \leq \sum_{k=0}^{\lfloor T/\eta \rfloor + 1} P_\gamma^k(a, B(0, (2\delta)^{-1}\varepsilon)^c),$$

where $B(0, r) \subset E$ stands for the ball of radius r and where P_γ^k stands for the iterated kernel, recursively defined by

$$P_\gamma^k(a, \cdot) = \int P_\gamma(a, dy) P_\gamma^{k-1}(y, \cdot) \quad (18)$$

and $P_\gamma^0(a, \cdot) = \delta_a$. Using (6), it is an easy exercise to show, by induction, that for every $k \in \mathbb{N}$, $P_\gamma^k(a, B(0, r)^c) \rightarrow 0$ as $r \rightarrow \infty$. By letting $\delta \rightarrow 0$ in the above inequality, the tightness of $(\mathbb{P}^{a_n, \gamma_n} \mathbf{X}_{\gamma_n}^{-1})$ follows.

5 Proof of Proposition 3.2

To establish Prop. 3.2-i), we consider a sequence $((\pi_n, \gamma_n))$ such that $\pi_n \in \mathcal{I}(P_{\gamma_n})$, $\gamma_n \rightarrow 0$, and (π_n) is tight. We first show that the sequence $(v_n := \mathbb{P}^{\pi_n, \gamma_n} \mathbf{X}_{\gamma_n}^{-1})$ is tight, then we show that every cluster point of (v_n) satisfies the conditions of Def. 2.1.

Given $\varepsilon > 0$, there exists a compact set $K \subset E$ such that $\inf_n \pi_n(K) > 1 - \varepsilon/2$. By Th. 3.1, the family $\{\mathbb{P}^{a, \gamma_n} \mathbf{X}_{\gamma_n}^{-1} : a \in K, n \in \mathbb{N}\}$ is tight. Let \mathcal{C} be a compact set of $C(\mathbb{R}_+, E)$ such that $\inf_{a \in K, n \in \mathbb{N}} \mathbb{P}^{a, \gamma_n} \mathbf{X}_{\gamma_n}^{-1}(\mathcal{C}) > 1 - \varepsilon/2$. By construction of the probability measure $\mathbb{P}^{\pi_n, \gamma_n}$, it holds that $\mathbb{P}^{\pi_n, \gamma_n}(\cdot) = \int_E \mathbb{P}^{a, \gamma_n}(\cdot) \pi_n(da)$. Thus,

$$v_n(\mathcal{C}) \geq \int_K \mathbb{P}^{a, \gamma_n} \mathbf{X}_{\gamma_n}^{-1}(\mathcal{C}) \pi_n(da) > (1 - \varepsilon/2)^2 > 1 - \varepsilon,$$

which shows that (v_n) is tight.

Since $\pi_n = v_n p_0^{-1}$, and since the projection p_0 is continuous, it is clear that every cluster point π of $\mathcal{I}(\mathcal{P})$ as $\gamma \rightarrow 0$ can be written as $\pi = v p_0^{-1}$, where v is a cluster point of a sequence (v_n) . Thus, Def. 2.1-(iii) is satisfied by π and v . To establish Prop. 3.2-i), we need to verify the conditions (i) and (ii) of Definition 2.1. In the remainder of the proof, we denote with a small abuse as (n) a subsequence along which (v_n) converges narrowly to v .

To establish the validity of Def. 2.1-(i), we prove that for every $\eta > 0$, $v_n((\Phi_H(E))_\eta) \rightarrow 1$ as $n \rightarrow \infty$; the result will follow from the convergence of (v_n) . Fix $\varepsilon > 0$, and let $K \subset E$ be a compact set such that $\inf_n \pi_n(K) > 1 - \varepsilon$. We have

$$\begin{aligned} v_n((\Phi_H(E))_\eta) &= \mathbb{P}^{\pi_n, \gamma_n}(d(\mathbf{X}_{\gamma_n}, \Phi_H(E)) < \eta) \\ &\geq \mathbb{P}^{\pi_n, \gamma_n}(d(\mathbf{X}_{\gamma_n}, \Phi_H(K)) < \eta) \\ &\geq \int_K \mathbb{P}^{a, \gamma_n}(d(\mathbf{X}_{\gamma_n}, \Phi_H(K)) < \eta) \pi_n(da) \\ &\geq (1 - \varepsilon) \inf_{a \in K} \mathbb{P}^{a, \gamma_n}(d(\mathbf{X}_{\gamma_n}, \Phi_H(K)) < \eta). \end{aligned}$$

By Th. 3.1, the infimum at the right hand side converges to 1. Since $\varepsilon > 0$ is arbitrary, we get the result.

It remains to establish the Θ -invariance of v (Condition (ii)). Equivalently, we need to show that

$$\int f(x) v(dx) = \int f \circ \Theta_t(x) v(dx) \quad (19)$$

for all $f \in C_b(C(\mathbb{R}_+, E))$ and all $t > 0$ (see [20] or [21, Th. 6.8]). We shall work on v_n and make $n \rightarrow \infty$. Write $\eta_n := t - \gamma_n \lfloor t/\gamma_n \rfloor$. Thanks to the P_{γ_n} -invariance of π_n , we have

$$\begin{aligned} \int f \circ \Theta_t(x) v_n(dx) &= \int f \circ \Theta_{\eta_n}(x(\gamma_n \lfloor t/\gamma_n \rfloor + \cdot)) v_n(dx) \\ &= \mathbb{P}^{\pi_n, \gamma_n} \mathbf{X}_{\gamma_n}^{-1}[f \circ \Theta_{\eta_n}(x(\gamma_n \lfloor t/\gamma_n \rfloor + \cdot))] \\ &= \mathbb{P}^{\pi_n, \gamma_n} \mathbf{X}_{\gamma_n}^{-1}[f \circ \Theta_{\eta_n}(x(\cdot))] \\ &= \int f \circ \Theta_{\eta_n}(x) v_n(dx). \end{aligned}$$

Denote as X^{π_n, γ_n} the canonical process on $E^{\mathbb{N}}$ endowed with the probability measure $\mathbb{P}^{\pi_n, \gamma_n}$, as defined in Sec. 3.1, and let $x^{\pi_n, \gamma_n} := X_{\gamma_n}(X^{\pi_n, \gamma_n})$. We showed above that the sequence (x^{π_n, γ_n}) is tight. Since $w_{\Theta_{\eta_n}(x)}^T(\delta) \leq w_x^{T+1}(\delta)$ for every $\delta, T > 0$ and for all n large, the sequence $(\Theta_{\eta_n}(x^{\pi_n, \gamma_n}))$

is also tight, thus, $(\Theta_{\eta_n}(\mathbf{x}^{\pi_n, \gamma_n}) - \mathbf{x}^{\pi_n, \gamma_n})$ is tight. Since $\eta_n \rightarrow 0$, we obtain by the continuity of $\mathbf{x}^{\pi_n, \gamma_n}$ that for all $t \geq 0$, $p_t(\Theta_{\eta_n}(\mathbf{x}^{\pi_n, \gamma_n}) - \mathbf{x}^{\pi_n, \gamma_n}) \rightarrow_n 0$. Therefore, $\Theta_{\eta_n}(\mathbf{x}^{\pi_n, \gamma_n}) - \mathbf{x}^{\pi_n, \gamma_n} \rightarrow_n 0$ in probability, where 0 denotes the zero function. Given two tight sequences of random variables (X_n) and (Y_n) on a metric space which satisfy $X_n - Y_n \rightarrow 0$ in probability, one can show that $\mathbb{E}f(X_n) - \mathbb{E}f(Y_n) \rightarrow 0$ for all continuous and bounded functions f . In our situation, this amounts to

$$\int f(\mathbf{x}) v_n(d\mathbf{x}) - \int f \circ \Theta_t(\mathbf{x}) v_n(d\mathbf{x}) \xrightarrow{n \rightarrow \infty} 0.$$

Since, in addition,

$$\int f(\mathbf{x}) v_n(d\mathbf{x}) \xrightarrow{n \rightarrow \infty} \int f(\mathbf{x}) v(d\mathbf{x}), \text{ and } \int f \circ \Theta_t(\mathbf{x}) v_n(d\mathbf{x}) \xrightarrow{n \rightarrow \infty} \int f \circ \Theta_t(\mathbf{x}) v(d\mathbf{x}),$$

the identity (19) holds true. Prop. 3.2-i) is proven.

We now prove Prop. 3.2-ii). Consider a sequence $((\mathbf{m}_n, \gamma_n))$ such that $\mathbf{m}_n \in \mathcal{I}(P_{\gamma_n})$, $\gamma_n \rightarrow 0$, and $\mathbf{m}_n \Rightarrow \mathbf{m}$ for some $\mathbf{m} \in \mathcal{M}(\mathcal{M}(E))$. Since the space $\mathcal{M}(E)$ is separable, Skorokhod's representation theorem shows that there exists a probability space $(\Omega', \mathcal{F}', \mathbb{P}')$, a sequence of $\Omega' \rightarrow \mathcal{M}(E)$ random variables (Λ_n) with distributions \mathbf{m}_n , and a $\Omega' \rightarrow \mathcal{M}(E)$ random variable Λ with distribution \mathbf{m} such that $\Lambda_n(\omega) \Rightarrow \Lambda(\omega)$ for each $\omega \in \Omega'$. Moreover, there is a probability one subset of Ω' such that $\Lambda_n(\omega)$ is a P_{γ_n} -invariant probability measure for all n and for every ω in this set. For each of these ω , we can construct on the space $(E^{\mathbb{N}}, \mathcal{F})$ a probability measure $\mathbb{P}^{\Lambda_n(\omega), \gamma_n}$ as we did in Sec. 3.1. By the same argument as in the proof of Prop. 3.2-i), the sequence $(\mathbb{P}^{\Lambda_n(\omega), \gamma_n} \mathbf{X}_{\gamma_n}^{-1})$ is tight, and any cluster point ν satisfies the conditions of Def. 2.1 with $\Lambda(\omega) = \nu p_0^{-1}$. Prop. 3.2 is proven.

6 Proof of Theorem 3.3

6.1 Technical lemmas

Lemma 6.1. Given a family $\{K_j : j \in \mathbb{N}\}$ of compact sets of E , the set

$$U := \{\nu \in \mathcal{M}(E) : \forall j \in \mathbb{N}, \nu(K_j) \geq 1 - 2^{-j}\}$$

is a compact set of $\mathcal{M}(E)$.

Proof. The set U is tight hence relatively compact by Prokhorov's theorem. It is moreover closed. Indeed, let (ν_n) represent a sequence of U s.t. $\nu_n \Rightarrow \nu$. Then, for all $j \in \mathbb{N}$, $\nu(K_j) \geq \limsup_n \nu_n(K_j) \geq 1 - 2^{-j}$ since K_j is closed. \square

Lemma 6.2. Let X be a real random variable such that $X \leq 1$ with probability one, and $\mathbb{E}X \geq 1 - \varepsilon$ for some $\varepsilon \geq 0$. Then $\mathbb{P}[X \geq 1 - \sqrt{\varepsilon}] \geq 1 - \sqrt{\varepsilon}$.

Proof. $1 - \varepsilon \leq \mathbb{E}X \leq \mathbb{E}X \mathbb{1}_{X < 1 - \sqrt{\varepsilon}} + \mathbb{E}X \mathbb{1}_{X \geq 1 - \sqrt{\varepsilon}} \leq (1 - \sqrt{\varepsilon})(1 - \mathbb{P}[X \geq 1 - \sqrt{\varepsilon}]) + \mathbb{P}[X \geq 1 - \sqrt{\varepsilon}]$. The result is obtained by rearranging. \square

For any $\mathbf{m} \in \mathcal{M}(\mathcal{M}(E))$, we denote by $e(\mathbf{m})$ the probability measure in $\mathcal{M}(E)$ such that for every $f \in C_b(E)$,

$$e(\mathbf{m}) : f \mapsto \int \nu(f) \mathbf{m}(d\nu).$$

Otherwise stated, $e(\mathbf{m})(f) = \mathbf{m}(\mathcal{T}_f)$ where $\mathcal{T}_f : \nu \mapsto \nu(f)$.

Lemma 6.3. Let \mathcal{L} be a family on $\mathcal{M}(\mathcal{M}(E))$. If $\{e(\mathbf{m}) : \mathbf{m} \in \mathcal{L}\}$ is tight, then \mathcal{L} is tight.

Proof. Let $\varepsilon > 0$ and choose any integer k s.t. $2^{-k+1} \leq \varepsilon$. For all $j \in \mathbb{N}$, choose a compact set $K_j \subset E$ s.t. for all $\mathbf{m} \in \mathcal{L}$, $e(\mathbf{m})(K_j) > 1 - 2^{-2j}$. Define U as the set of measures $\nu \in \mathcal{M}(E)$ s.t. for all $j \geq k$, $\nu(K_j) \geq 1 - 2^{-j}$. By Lemma 6.1, U is compact. For all $\mathbf{m} \in \mathcal{L}$, the union bound implies that

$$\mathbf{m}(E \setminus U) \leq \sum_{j=k}^{\infty} \mathbf{m}\{\nu : \nu(K_j) < 1 - 2^{-j}\}$$

By Lemma 6.2, $\mathbf{m}\{\nu : \nu(K_j) \geq 1 - 2^{-j}\} \geq 1 - 2^{-j}$. Therefore, $\mathbf{m}(E \setminus U) \leq \sum_{j=k}^{\infty} 2^{-j} = 2^{-k+1} \leq \varepsilon$. This proves that \mathcal{L} is tight. \square

Lemma 6.4. Let $(\mathbf{m}_n : n \in \mathbb{N})$ be a sequence on $\mathcal{M}(\mathcal{M}(E))$, and consider $\bar{\mathbf{m}} \in \mathcal{M}(\mathcal{M}(E))$. If $\mathbf{m}_n \Rightarrow \bar{\mathbf{m}}$, then $e(\mathbf{m}_n) \Rightarrow e(\bar{\mathbf{m}})$.

Proof. For any $f \in C_b(E)$, $\mathcal{T}_f \in C_b(\mathcal{M}(E))$. Thus, $\mathbf{m}_n(\mathcal{T}_f) \rightarrow \bar{\mathbf{m}}(\mathcal{T}_f)$. \square

When a sequence \mathbf{m}_n of $\mathcal{M}(\mathcal{M}(E))$ converges narrowly to $\mathbf{m} \in \mathcal{M}(\mathcal{M}(E))$, it follows from the above proof that $\mathbf{m}_n \mathcal{T}_f^{-1} \Rightarrow \mathbf{m} \mathcal{T}_f^{-1}$ for all bounded continuous f . The purpose of the next lemma is to extend this result to the case where f is not necessarily bounded, but instead, satisfies some uniform integrability condition. For any vector-valued function f , we use the notation $\|f\| := \|f(\cdot)\|$.

Lemma 6.5. Let $f \in C(E, \mathbb{R}^{N'})$ where $N' \geq 1$ is an integer. Define by $\mathcal{T}_f : \mathcal{M}(E) \rightarrow \mathbb{R}$ the mapping s.t. $\mathcal{T}_f(\nu) := \nu(f)$ if $\nu(\|f\|) < \infty$ and equal to zero otherwise. Let $(\mathbf{m}_n : n \in \mathbb{N})$ be a sequence on $\mathcal{M}(\mathcal{M}(E))$ and let $\mathbf{m} \in \mathcal{M}(\mathcal{M}(E))$. Assume that $\mathbf{m}_n \Rightarrow \mathbf{m}$ and

$$\lim_{K \rightarrow \infty} \sup_n e(\mathbf{m}_n)(\|f\| \mathbb{1}_{\|f\| > K}) = 0. \quad (20)$$

Then, $\nu(\|f\|) < \infty$ for all ν \mathbf{m} -a.e. and $\mathbf{m}_n \mathcal{T}_f^{-1} \Rightarrow \mathbf{m} \mathcal{T}_f^{-1}$.

Proof. By Eq. (20), $e(\mathbf{m})(\|f\|) < \infty$. This implies that for all ν \mathbf{m} -a.e., $\nu(\|f\|) < \infty$. Choose $h \in C_b(\mathbb{R}^{N'})$ s.t. h is L -Lipschitz continuous. We must prove that $\mathbf{m}_n \mathcal{T}_f^{-1}(h) \rightarrow \mathbf{m} \mathcal{T}_f^{-1}(h)$. By the above remark, $\mathbf{m} \mathcal{T}_f^{-1}(h) = \int h(\nu(f)) d\mathbf{m}(\nu)$, and by Eq (20), $\mathbf{m}_n \mathcal{T}_f^{-1}(h) = \int h(\nu(f)) d\mathbf{m}_n(\nu)$. Choose $\varepsilon > 0$. By Eq. (20), there exists $K_0 > 0$ s.t. for all $K > K_0$, $\sup_n e(\mathbf{m}_n)(\|f\| \mathbb{1}_{\|f\| > K}) < \varepsilon$. For every such K , define the bounded function $f_K \in C(E, \mathbb{R}^{N'})$ by $f_K(x) = f(x)(1 \wedge K/\|f(x)\|)$. For all $K > K_0$, and for all $n \in \mathbb{N}$,

$$\begin{aligned} |\mathbf{m}_n \mathcal{T}_f^{-1}(h) - \mathbf{m}_n \mathcal{T}_{f_K}^{-1}(h)| &\leq \int |h(\nu(f)) - h(\nu(f_K))| d\mathbf{m}_n(\nu) \\ &\leq L \int \nu(\|f - f_K\|) d\mathbf{m}_n(\nu) \\ &\leq L \int \nu(\|f\| \mathbb{1}_{\|f\| > K}) d\mathbf{m}_n(\nu) \leq L\varepsilon. \end{aligned}$$

By continuity of \mathcal{T}_{f_K} , it holds that $\mathbf{m}_n \mathcal{T}_{f_K}^{-1}(h) \rightarrow \mathbf{m} \mathcal{T}_{f_K}^{-1}(h)$. Therefore, for every $K > K_0$, $\limsup_n |\mathbf{m}_n \mathcal{T}_f^{-1}(h) - \mathbf{m} \mathcal{T}_f^{-1}(h)| \leq L\varepsilon$. As $\nu(\|f\|) < \infty$ for all ν \mathbf{m} -a.e., the dominated convergence theorem implies that $\nu(f_K) \rightarrow \nu(f)$ as $K \rightarrow \infty$, \mathbf{m} -a.e. As h is bounded and continuous, a second application of the dominated convergence theorem implies that $\int h(\nu(f_K)) d\mathbf{m}(\nu) \rightarrow \int h(\nu(f)) d\mathbf{m}(\nu)$, which reads $\mathbf{m} \mathcal{T}_{f_K}^{-1}(h) \rightarrow \mathbf{m} \mathcal{T}_f^{-1}(h)$. Thus, $\limsup_n |\mathbf{m}_n \mathcal{T}_f^{-1}(h) - \mathbf{m} \mathcal{T}_f^{-1}(h)| \leq L\varepsilon$. As a consequence, $\mathbf{m}_n \mathcal{T}_f^{-1}(h) \rightarrow \mathbf{m} \mathcal{T}_f^{-1}(h)$ as $n \rightarrow \infty$, which completes the proof. \square

6.2 Narrow Cluster Points of the Empirical Measures

Let $P : E \times \mathcal{B}(E) \rightarrow [0, 1]$ be a probability transition kernel. For $\nu \in \mathcal{M}(E)$, we denote by $\mathbb{P}^{\nu, P}$ the probability on (Ω, \mathcal{F}) such that X is an homogeneous Markov chain with initial distribution ν and transition kernel P .

For every $n \in \mathbb{N}$, we define the measurable mapping $\Lambda_n : \Omega \rightarrow \mathcal{M}(E)$ as

$$\Lambda_n(x) := \frac{1}{n+1} \sum_{k=0}^n \delta_{x_k} \quad (21)$$

for all $x = (x_k : k \in \mathbb{N})$. Note that

$$\mathbb{E}^{\nu, P} \Lambda_n = \frac{1}{n+1} \sum_{k=0}^n \nu P^k,$$

where P^k stands for the iterated kernel, recursively defined by $P^k(x, \cdot) = \int P(x, dy) P^{k-1}(y, \cdot)$ and $P^0(x, \cdot) = \delta_x$.

We recall that $\mathcal{J}(P)$ represents the subset of $\mathcal{M}(\mathcal{M}(E))$ formed by the measures whose support is included in $\mathcal{I}(P)$.

Proposition 6.6. Let $P : E \times \mathcal{B}(E) \rightarrow [0, 1]$ be a Feller probability transition kernel. Let $\nu \in \mathcal{M}(E)$.

1. Any cluster point of $\{\mathbb{E}^{\nu, P} \Lambda_n : n \in \mathbb{N}\}$ is an element of $\mathcal{I}(P)$.
2. Any cluster point of $\{\mathbb{P}^{\nu, P} \Lambda_n^{-1} : n \in \mathbb{N}\}$ is an element of $\mathcal{J}(P)$.

Proof. We omit the upper script ν, P . For all $f \in C_b(E)$, $\mathbb{E} \Lambda_n(Pf) - \mathbb{E} \Lambda_n(f) \rightarrow 0$. As P is Feller, any cluster point π of $\{\mathbb{E} \Lambda_n : n \in \mathbb{N}\}$ satisfies $\pi(Pf) = \pi(f)$. This proves the first point.

For every $f \in C_b(E)$ and $x \in \Omega$, consider the decomposition:

$$\Lambda_n(x)(Pf) - \Lambda_n(x)(f) = \frac{1}{n+1} \sum_{k=0}^{n-1} (Pf(x_k) - f(x_{k+1})) + \frac{Pf(x_n) - f(x_0)}{n+1}.$$

Using that f is bounded, Doob's martingale convergence theorem implies that the sequence $\left(\sum_{k=0}^{n-1} k^{-1} (Pf(X_k) - f(X_{k+1})) \right)_n$ converges a.s. when n tends to infinity. By Kronecker's lemma, we deduce that $\frac{1}{n+1} \sum_{k=0}^{n-1} (Pf(X_k) - f(X_{k+1}))$ tends a.s. to zero. Hence,

$$\Lambda_n(Pf) - \Lambda_n(f) \rightarrow 0 \text{ a.s.} \quad (22)$$

Now consider a subsequence (Λ_{φ_n}) which converges in distribution to some r.v. Λ as n tends to infinity. For a fixed $f \in C_b(E)$, the mapping $\nu \mapsto (\nu(f), \nu(Pf))$ on $\mathcal{M}(\mathbb{R}) \rightarrow \mathbb{R}^2$ is continuous. From the mapping theorem, $\Lambda_{\varphi_n}(f) - \Lambda_{\varphi_n}(Pf)$ converges in distribution to $\Lambda(f) - \Lambda(Pf)$. By (22), it follows that $\Lambda(f) - \Lambda(Pf) = 0$ on some event $\mathcal{E}_f \in \mathcal{F}$ of probability one. Denote by $C_\kappa(E) \subset C_b(E)$ the set of continuous real-valued functions having a compact support, and let $C_\kappa(E)$ be equipped with the uniform norm $\|\cdot\|_\infty$. Introduce a dense denumerable subset S of $C_\kappa(E)$. On the probability-one event $\mathcal{E} = \cap_{f \in D} \mathcal{E}_f$, it holds that for all $f \in S$, $\Lambda(f) = \Lambda(Pf)$. Now consider $g \in C_\kappa(E)$ and let $\varepsilon > 0$. Choose $f \in S$ such that $\|f - g\|_\infty \leq \varepsilon$. Then, almost everywhere on \mathcal{E} , $|\Lambda(g) - \Lambda(Pg)| \leq |\Lambda(f) - \Lambda(g)| + |\Lambda(Pf) - \Lambda(Pg)| \leq 2\varepsilon$. Thus, $\Lambda(g) - \Lambda(Pg) = 0$ for every $g \in C_\kappa(E)$. Hence, almost everywhere on \mathcal{E} , one has $\Lambda = \Lambda P$. \square

6.3 Tightness of the Empirical Measures

Proposition 6.7. Let \mathcal{P} be a family of transition kernels on E . Let $V : E \rightarrow [0, +\infty)$, $\psi : E \rightarrow [0, +\infty)$ be measurable. Let $\alpha : \mathcal{P} \rightarrow (0, +\infty)$, $\beta : \mathcal{P} \rightarrow \mathbb{R}$. Assume that $\sup_{P \in \mathcal{P}} \frac{\beta(P)}{\alpha(P)} < \infty$, $\sup_{P \in \mathcal{P}} \frac{\beta(P)}{\alpha(P)} < \infty$ and assume that for every $P \in \mathcal{P}$,

$$PV \leq V - \alpha(P)\psi + \beta(P).$$

Then, the following holds.

- i) The family $\bigcup_{P \in \mathcal{P}} \mathcal{I}(P)$ is tight. Moreover, $\sup_{\pi \in \mathcal{I}(\mathcal{P})} \pi(\psi) < +\infty$.
- ii) For every $\nu \in \mathcal{M}(E)$ s.t. $\nu(V) < \infty$, every $P \in \mathcal{P}$, $\{\mathbb{E}^{\nu, P} \Lambda_n : n \in \mathbb{N}\}$ is tight. Moreover, $\sup_{n \in \mathbb{N}} \mathbb{E}^{\nu, P} \Lambda_n(\psi) < \infty$.

Proof. For each $P \in \mathcal{P}$, PV is everywhere finite by Condition (PH). Moreover,

$$\sum_{k=0}^n P^{k+1}V \leq \sum_{k=0}^n P^kV - \alpha(P) \sum_{k=0}^n P^k\psi + (n+1)\beta(P).$$

Using that $V \geq 0$ and $\alpha(P) > 0$,

$$\frac{1}{n+1} \sum_{k=0}^n P^k\psi \leq \frac{V}{\alpha(P)(n+1)} + c,$$

where $c := \sup_{P \in \mathcal{P}} \beta(P)/\alpha(P)$ is finite. For any $M > 0$,

$$\begin{aligned} \frac{1}{n+1} \sum_{k=0}^n P^k(\psi \wedge M) &\leq \left(\frac{1}{n+1} \sum_{k=0}^n P^k\psi \right) \wedge M \\ &\leq \left(\frac{V}{\alpha(P)(n+1)} + c \right) \wedge M. \end{aligned} \tag{23}$$

Set $\pi \in \mathcal{I}(\mathcal{P})$, and consider $P \in \mathcal{P}$ such that $\pi = \pi P$. Inequality (23) implies that for every n ,

$$\pi(\psi \wedge M) \leq \pi \left(\left(\frac{V}{\alpha(P)(n+1)} + c \right) \wedge M \right).$$

By Lebesgue's dominated convergence theorem, $\pi(\psi \wedge M) \leq c$. Letting $M \rightarrow \infty$ yields $\pi(\psi) \leq c$. The tightness of $\mathcal{I}(\mathcal{P})$ follows from the coercivity of ψ . Setting $M = +\infty$ in (23), and integrating w.r.t. ν , we obtain

$$\mathbb{E}^{\nu, P} \Lambda_n(\psi) \leq \frac{\nu(V)}{(n+1)\alpha(P)} + c,$$

which proves the second point. \square

Proposition 6.8. We posit the assumptions of Prop. 6.7. Then,

- 1. The family $\mathcal{J}(\mathcal{P}) := \bigcup_{P \in \mathcal{P}} \mathcal{J}(P)$ is tight;
- 2. $\{\mathbb{P}^{\nu, P} \Lambda_n^{-1} : n \in \mathbb{N}\}$ is tight.

Proof. For every $\mathbf{m} \in \mathcal{J}(\mathcal{P})$, it is easy to see that $e(\mathbf{m}) \in \mathcal{I}(\mathcal{P})$. Thus, $\{e(\mathbf{m}) : \mathbf{m} \in \mathcal{J}(\mathcal{P})\}$ is tight by Prop. 6.7. By Lemma 6.3, $\mathcal{J}(\mathcal{P})$ is tight. The second point follows from the equality $\mathbb{E}^{\nu, P} \Lambda_n = e(\mathbb{P}^{\nu, P} \Lambda_n^{-1})$ along with Prop. 6.7 and Lemma 6.4. \square

6.4 Main Proof

By continuity of $h_\gamma(s, \cdot)$ for every $s \in \Xi$, $\gamma \in (0, \gamma_0)$, the transition kernel P_γ given by (5) is Feller. By Prop. 6.7 and Eq. (8), we have $\sup_n \mathbb{E}^{\nu, \gamma} \Lambda_n(\varphi \circ f) < \infty$ which, by de la Vallée-Poussin's criterion for uniform integrability, implies

$$\lim_{K \rightarrow \infty} \sup_n \mathbb{E}^{\nu, \gamma} \Lambda_n(\|f\| \mathbb{1}_{\|f\| > K}) = 0. \tag{24}$$

In particular, the quantity $\mathbb{E}^{\nu, \gamma} \Lambda_n(f) = \mathbb{E}^{\nu, \gamma}(F_n)$ is well-defined.

We now prove the statement (9). By contradiction, assume that for some $\delta > 0$, there exists a positive sequence $\gamma_j \rightarrow 0$, such that for all $j \in \mathbb{N}$, $\limsup_{n \rightarrow \infty} d(\mathbb{E}^{\nu, \gamma_j} \Lambda_n(f), \mathcal{S}_f) > \delta$. For every j , there exists an increasing sequence of integers $(\varphi_n^j : n \in \mathbb{N})$ converging to $+\infty$ s.t.

$$\forall n, d(\mathbb{E}^{\nu, \gamma_j} \Lambda_{\varphi_n^j}(f), \mathcal{S}_f) > \delta. \quad (25)$$

By Prop. 6.7, the sequence $(\mathbb{E}^{\nu, \gamma_j} \Lambda_{\varphi_n^j} : n \in \mathbb{N})$ is tight. By Prokhorov's theorem and Prop. 6.6, there exists $\pi_j \in \mathcal{I}(P_{\gamma_j})$ such that, as n tends to infinity, $\mathbb{E}^{\nu, \gamma_j} \Lambda_{\varphi_n^j} \Rightarrow \pi_j$ along some subsequence. By the uniform integrability condition (24), $\pi_j(\|f\|) < \infty$ and $\mathbb{E}^{\nu, \gamma_j} \Lambda_{\varphi_n^j}(f) \rightarrow \pi_j(f)$ as n tends to infinity, along the latter subsequence. By Eq. (25), for all $j \in \mathbb{N}$, $d(\pi_j(f), \mathcal{S}_f) > \delta$. By Prop. 6.7, $\sup_{\pi \in \mathcal{I}(\mathcal{P})} \pi(\psi) < +\infty$. Since $\varphi \circ f \leq M(1 + \psi)$, de la Vallée-Poussin's criterion again implies that

$$\sup_{\pi \in \mathcal{I}(\mathcal{P})} \pi(\|f\| \mathbb{1}_{\|f\| > K}) < \infty. \quad (26)$$

Also by Prop. 6.7, the sequence (π_j) is tight. Thus $\pi_j \Rightarrow \pi$ along some subsequence, for some measure π which, by Prop. 3.2, is invariant for Φ_H . The uniform integrability condition (26) implies that $\pi(\|f\|) < \infty$ (hence, the set \mathcal{S}_f is non-empty) and $\pi_j(f) \rightarrow \pi(f)$ as j tends to infinity, along the above subsequence. This shows that $d(\pi(f), \mathcal{S}_f) > \delta$, which is absurd. The statement (9) holds true (and in particular, \mathcal{S}_f must be non-empty).

The proof of the statement (7) follows the same line, by replacing f with the function $\mathbb{1}_{\overline{\mathcal{U}_\epsilon^c}}$. We briefly explain how the proof adapts, without repeating all the arguments. In this case, $\mathcal{S}_{\mathbb{1}_{\mathcal{U}_\epsilon^c}}$ is the singleton $\{0\}$, and Equation (25) reads $\mathbb{E}^{\nu, \gamma_j} \Lambda_{\varphi_n^j}(\mathcal{U}_\epsilon^c) > \delta$. By the Portmanteau theorem, $\limsup_n \mathbb{E}^{\nu, \gamma_j} \Lambda_{\varphi_n^j}(\mathcal{U}_\epsilon^c) \leq \pi_j(\mathcal{U}_\epsilon^c)$ where the limsup is taken along some subsequence. The contradiction follows from the fact that $\limsup \pi_j(\mathcal{U}_\epsilon^c) \leq \pi(\overline{\mathcal{U}_\epsilon^c}) = 0$ (where the limsup is again taken along the relevant subsequence).

We prove the statement (10). Assume by contradiction that for some (other) sequence $\gamma_j \rightarrow 0$, $\limsup_{n \rightarrow \infty} \mathbb{P}^{\nu, \gamma_j}(d(\Lambda_n(f), \mathcal{S}_f) \geq \varepsilon) > \delta$. For every j , there exists a sequence $(\varphi_n^j : n \in \mathbb{N})$ s.t.

$$\forall n, \mathbb{P}^{\nu, \gamma_j}(d(\Lambda_{\varphi_n^j}(f), \mathcal{S}_f) \geq \varepsilon) > \delta. \quad (27)$$

By Prop. 6.8, $(\mathbb{P}^{\nu, \gamma_j} \Lambda_{\varphi_n^j}^{-1} : n \in \mathbb{N})$ is tight, one can extract a further subsequence (which we still denote by (φ_n^j) for simplicity) s.t. $\mathbb{P}^{\nu, \gamma_j} \Lambda_{\varphi_n^j}^{-1}$ converges narrowly to a measure \mathbf{m}_j as n tends to infinity, which, by Prop. 6.6, satisfies $\mathbf{m}_j \in \mathcal{J}(P_{\gamma_j})$. Noting that $e(\mathbb{P}^{\nu, \gamma_j} \Lambda_{\varphi_n^j}^{-1}) = \mathbb{E}^{\nu, \gamma_j} \Lambda_{\varphi_n^j}$ and recalling Eq. (24), Lemma 6.5 implies that $\nu'(\|f\|) < \infty$ for all ν' \mathbf{m}_j -a.e., and $\mathbb{P}^{\nu, \gamma_j} \Lambda_{\varphi_n^j}^{-1} \mathcal{T}_f \Rightarrow \mathbf{m}_j \mathcal{T}_f^{-1}$, where we recall that $\mathcal{T}_f(\nu') := \nu'(f)$ for all ν' s.t. $\nu'(\|f\|) < \infty$. As $(\mathcal{S}_f)_\varepsilon^c$ is a closed set,

$$\begin{aligned} \mathbf{m}_j \mathcal{T}_f^{-1}((\mathcal{S}_f)_\varepsilon^c) &\geq \limsup_n \mathbb{P}^{\nu, \gamma_j} \Lambda_{\varphi_n^j}^{-1} \mathcal{T}_f^{-1}((\mathcal{S}_f)_\varepsilon^c) \\ &= \limsup_n \mathbb{P}^{\nu, \gamma_j}(d(\Lambda_{\varphi_n^j}(f), \mathcal{S}_f) \geq \varepsilon) > \delta. \end{aligned}$$

By Prop. 6.7, (\mathbf{m}_j) is tight, and one can extract a subsequence (still denoted by (\mathbf{m}_j)) along which $\mathbf{m}_j \Rightarrow \mathbf{m}$ for some measure \mathbf{m} which, by Prop. 3.2, belongs to $\mathcal{J}(\Phi_H)$. For every j , $e(\mathbf{m}_j) \in \mathcal{I}(P_{\gamma_j})$. By the uniform integrability condition (26), one can apply Lemma 6.5 to the sequence (\mathbf{m}_j) . We deduce that $\nu'(\|f\|) < \infty$ for all ν' \mathbf{m} -a.e. and $\mathbf{m}_j \mathcal{T}_f^{-1} \Rightarrow \mathbf{m} \mathcal{T}_f^{-1}$. In particular,

$$\mathbf{m} \mathcal{T}_f^{-1}((\mathcal{S}_f)_\varepsilon^c) \geq \limsup_j \mathbf{m}_j \mathcal{T}_f^{-1}((\mathcal{S}_f)_\varepsilon^c) > \delta.$$

Since $\mathbf{m} \in \mathcal{J}(\Phi_H)$, it holds that $\mathbf{m} \mathcal{T}_f^{-1}((\mathcal{S}_f)_\varepsilon^c) = 0$, hence a contradiction.

7 Proofs of Theorems 3.4 and 3.5

7.1 Proof of Theorem 3.4

In this proof, we set $L = L_{\text{av}(\Phi)}$ to simplify the notations. It is straightforward to show that the identity mapping $f(x) = x$ satisfies the hypotheses of Th. 3.3. Hence, it is sufficient to prove that \mathcal{S}_f is a subset of $\overline{\text{co}}(L)$, the closed convex hull of L . Choose $q \in \mathcal{S}_f$ and let $q = \int x d\pi(x)$ for some $\pi \in \mathcal{I}(\Phi)$ admitting a first order moment. There exists a Θ -invariant measure $v \in \mathcal{M}(C(\mathbb{R}_+, E))$ s.t. $\text{supp}(v) \subset \Phi(E)$ and $vp_0^{-1} = \pi$. We remark that for all $t > 0$,

$$q = v(p_0) = v(p_t) = v(p_t \circ \text{av}), \quad (28)$$

where the second identity is due to the shift-invariance of v , and the last one uses Fubini's theorem. Again by the shift-invariance of v , the family $\{p_t : t > 0\}$ is uniformly integrable w.r.t. v . By Tonelli's theorem, $\sup_{t>0} v(\|p_t \circ \text{av}\| \mathbb{1}_S) \leq \sup_{t>0} v(\|p_t\| \mathbb{1}_S)$ for every $S \in \mathcal{B}(C(\mathbb{R}_+, E))$. Hence, the family $\{p_t \circ \text{av} : t > 0\}$ is v -uniformly integrable as well. In particular, $\{p_t \circ \text{av} : t > 0\}$ is tight in $(C(\mathbb{R}_+, E), \mathcal{B}(C(\mathbb{R}_+, E)), v)$. By Prokhorov's theorem, there exists a sequence $t_n \rightarrow \infty$ and a measurable function $g : C(\mathbb{R}_+, E) \rightarrow E$ such that $p_{t_n} \circ \text{av}$ converges in distribution to g as $n \rightarrow \infty$. By uniform integrability, $v(p_{t_n} \circ \text{av}) \rightarrow v(g)$. Equation (28) finally implies that

$$q = v(g).$$

In order to complete the proof, it is sufficient to show that $g(x) \in \overline{L}$ for every x v -a.e., because $\overline{\text{co}}(L) \subset \text{co}(\overline{L})$. Set $\varepsilon > 0$ and $\delta > 0$. By the tightness of the r.v. $(p_{t_n} \circ \text{av} : n \in \mathbb{N})$, choose a compact set K such that $v(p_{t_n} \circ \text{av})^{-1}(K^c) \leq \delta$ for all n . As $\overline{L_\varepsilon^c}$ is an open set, one has

$$vg^{-1}(\overline{L_\varepsilon^c}) \leq \lim_n v(p_{t_n} \circ \text{av})^{-1}(\overline{L_\varepsilon^c}) \leq \lim_n v(p_{t_n} \circ \text{av})^{-1}(\overline{L_\varepsilon^c} \cap K) + \delta.$$

Let $x \in \Phi(E)$ be fixed. By contradiction, suppose that $\mathbb{1}_{\overline{L_\varepsilon^c} \cap K}(p_{t_n}(\text{av}(x)))$ does not converge to zero. Then, $p_{t_n}(\text{av}(x)) \in \overline{L_\varepsilon^c} \cap K$ for every n along some subsequence. As K is compact, one extract a subsequence, still denoted by t_n , s.t. $p_{t_n}(\text{av}(x))$ converges. The corresponding limit must belong to the closed set $\overline{L_\varepsilon^c}$, but must also belong to L by definition of x . This proves that $\mathbb{1}_{\overline{L_\varepsilon^c} \cap K}(p_{t_n} \circ \text{av}(x))$ converges to zero for all $x \in \Phi(E)$. As $\text{supp}(v) \subset \Phi(E)$, $\mathbb{1}_{\overline{L_\varepsilon^c} \cap K}(p_{t_n} \circ \text{av})$ converges to zero v -a.s. By the dominated convergence theorem, we obtain that $vg^{-1}(\overline{L_\varepsilon^c}) \leq \delta$. Letting $\delta \rightarrow 0$ we obtain that $vg^{-1}(\overline{L_\varepsilon^c}) = 0$. Hence, $g(x) \in \overline{L}$ for all x v -a.e. The proof is complete.

7.2 Proof of Theorem 3.5

Recall the definition $\mathcal{U} := \bigcup_{\pi \in \mathcal{I}(\Phi)} \text{supp}(\pi)$. By Th. 3.3, for all $\varepsilon > 0$,

$$\limsup_{n \rightarrow \infty} \mathbb{E}^{\nu, \gamma} \Lambda_n(\mathcal{U}_\varepsilon^c) \xrightarrow{\gamma \rightarrow 0} 0,$$

where Λ_n is the random measure given by (21). By Theorem 2.1, $\text{supp}(\pi) \subset \text{BC}_\Phi$ for each $\pi \in \mathcal{I}(\Phi)$. Thus, $\mathcal{U}_\varepsilon \subset (\text{BC}_\Phi)_\varepsilon$. Hence, $\limsup_n \mathbb{E}^{\nu, \gamma} \Lambda_n((\text{BC}_\Phi)_\varepsilon^c) \rightarrow 0$ as $\gamma \rightarrow 0$. This completes the proof.

References

- [1] J.-P. Aubin and A. Cellina. *Differential inclusions*, volume 264 of *Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences]*. Springer-Verlag, Berlin, 1984. Set-valued maps and viability theory.
- [2] J.-P. Aubin, H. Frankowska, and A. Lasota. Poincaré's recurrence theorem for set-valued dynamical systems. *Ann. Polon. Math.*, 54(1):85–91, 1991.

- [3] M. Benaïm. Dynamics of stochastic approximation algorithms. In *Séminaire de Probabilités, XXXIII*, volume 1709 of *Lecture Notes in Math.*, pages 1–68. Springer, Berlin, 1999.
- [4] M. Benaïm and M. W. Hirsch. Stochastic approximation algorithms with constant step size whose average is cooperative. *Ann. Appl. Probab.*, 9(1):216–241, 1999.
- [5] M. Benaïm, J. Hofbauer, and S. Sorin. Stochastic approximations and differential inclusions. *SIAM J. Control Optim.*, 44(1):328–348 (electronic), 2005.
- [6] M. Benaïm, J. Hofbauer, and S. Sorin. Stochastic approximations and differential inclusions. II. Applications. *Math. Oper. Res.*, 31(4):673–695, 2006.
- [7] A. Benveniste, M. Métivier, and P. Priouret. *Adaptive algorithms and stochastic approximations*, volume 22 of *Applications of Mathematics (New York)*. Springer-Verlag, Berlin, 1990. Translated from the French by Stephen S. Wilson.
- [8] P. Bianchi. Ergodic convergence of a stochastic proximal point algorithm. *SIAM J. Optim.*, 26(4):2235–2260, 2016.
- [9] P. Bianchi and W. Hachem. Dynamical behavior of a stochastic forward-backward algorithm using random monotone operators. *J. Optim. Theory Appl.*, 171(1):90–120, 2016.
- [10] P. Billingsley. *Convergence of probability measures*. Wiley Series in Probability and Statistics: Probability and Statistics. John Wiley & Sons, Inc., New York, second edition, 1999. A Wiley-Interscience Publication.
- [11] V. S. Borkar. *Stochastic approximation. A dynamical systems viewpoint*. Cambridge University Press, Cambridge; Hindustan Book Agency, New Delhi, 2008.
- [12] M. Faure and G. Roth. Stochastic approximations of set-valued dynamical systems: convergence with positive probability to an attractor. *Math. Oper. Res.*, 35(3):624–640, 2010.
- [13] M. Faure and G. Roth. Ergodic properties of weak asymptotic pseudotrajectories for set-valued dynamical systems. *Stoch. Dyn.*, 13(1):1250011, 23, 2013.
- [14] J.-C. Fort and G. Pagès. Asymptotic behavior of a Markovian stochastic algorithm with constant step. *SIAM J. Control Optim.*, 37(5):1456–1482 (electronic), 1999.
- [15] N. Gast and B. Gaujal. Markov chains with discontinuous drifts have differential inclusion limits. *Performance Evaluation*, 69(12):623 – 642, 2012.
- [16] R.Z. Ha’sminskii. The average principle for parabolic and elliptic differential equations and Markov processes with small diffusions. *Theor. Probab. Appl.*, 8:1–21, 1963.
- [17] H. J. Kushner and G. G. Yin. *Stochastic approximation and recursive algorithms and applications*, volume 35 of *Applications of Mathematics (New York)*. Springer-Verlag, New York, second edition, 2003. Stochastic Modelling and Applied Probability.
- [18] W. Miller and E. Akin. Invariant measures for set-valued dynamical systems. *Trans. Amer. Math. Soc.*, 351(3):1203–1225, 1999.
- [19] I. Molchanov. *Theory of random sets*. Springer Science & Business Media, 2006.
- [20] G. Roth and W. H. Sandholm. Stochastic approximations with constant step size and differential inclusions. *SIAM J. Control Optim.*, 51(1):525–555, 2013.
- [21] P. Walters. *An introduction to ergodic theory*, volume 79 of *Graduate Texts in Mathematics*. Springer-Verlag, New York-Berlin, 1982.
- [22] V. Yaji and S. Bhatnagar. Stochastic recursive inclusions with non-additive iterate-dependent Markov noise. *arXiv preprint arXiv:1607.04735*, 2016.